7 DIAGNOSIS AND RECOMMENDATIONS

> The high Phase III negative outcome rate problem [of clinical trials] is endemic, affecting sepsis, stroke, cancer, cardiology and orthopaedics research, to name just a few. These persistent failures have had a chilling effect on pharmaceutical industry investments in new drug development and the costs to pharma and government are staggering. It is no wonder that funding agencies and policy-makers in both sectors are deeply concerned and realize that continuing to do the same things in the same way cannot go on. Despite much recent discussion, little has been done to change the situation for the better. This paralysis could be due in part to the lack of any broad consensus about where the most basic problems lie.
>
> —STEIN (2015), P. 1259

Although biomedical research has made enormous progress over the last 200 years, alleviating a great deal of human suffering, it does have a problem. As stated in the chapter's opening quotation, far too many therapies show promise in animal and in vitro experiments but fail in clinical trials. Findings in toxicology also do not translate from animal and in vitro systems to humans as reliably as one would hope. The human, animal, and financial costs of these failures are high, but charting a more successful path is difficult. Different people are likely to offer different remedies, ranging from calls for increased rigor in preclinical research to championing experiments on human volunteers, rather than animals. I, too, feel torn between the various options. The reduction of animal suffering is certainly a worthy goal, especially if animals are dying in vain; yet human suffering is also a significant concern. If we do accept the need for animal research, should that research focus on species that are most similar to us, given that those animals are also most likely to feel and think the way we do? If we opt for studying more distant relatives, or even in vitro systems, are we inviting more failures of translation, thus wasting time, resources, and lives?

I struggled mightily with these questions as I was working on this book but cannot, in the end, offer a simple, clear remedy. Instead, I have opted for giving you, the reader, a trove of information and ideas that seem relevant to the problem of translational failure and success. Overall, I hope the book will help you think this problem through and make well-informed, deliberate decisions about your own personal way forward. If you are in a position of influence, this book will hopefully help you appreciate the problem's nuances and make wise decisions on policy. With those overarching goals in mind, I use this last chapter to present four very different perspectives on the use of material models in biology, followed by my own attempt at synthesis and some specific recommendations.

7.1 FOUR PERSPECTIVES ON MODELS IN BIOLOGY

How should researchers use animal and cellular models to develop effective treatments for human disease? Given that progress in this area has been slower than we want, do we need a radical course correction, or are we simply going through a spell of mediocre luck? As the opening quotation asks, where do the most basic problems lie, and how can they be fixed? When we try to answer these questions, each of us is likely to come at the problem from a somewhat different perspective, each addressing only part of the problem. We are like the proverbial blind men examining an elephant, each coming to quite different conclusions. In the following sections, I sketch four of these disparate perspectives and briefly critique them. Although my descriptions are certainly strawmen cartoons, delineating these viewpoints should, I hope, help us appreciate the problem's full complexity.

7.1.1 The Animal Welfare Perspective

In this book, we discussed many instances in which data from animal research failed to predict a compound's safety or efficacy in humans. A paradigm example is the trial of TGN1412, which had passed safety tests in rodents and monkeys but caused life-threatening harm in human volunteers (see chapter 1). One response to such translational failures is to argue that animal experiments in general cannot be trusted and should, therefore, be stopped (Shanks et al., 2009; Pound & Bracken, 2014). A milder, more widely accepted version of this anti-vivisectionist stance is to argue that researchers should reduce the number of animals they use and minimize their suffering. In particular, they should experiment on presumably less sentient animals—that is, animals thought to be not very capable of having feelings (Harnad, 2016)—or insentient materials whenever feasible. I call this the animal welfarist's perspective.

Complete replacement of sentient animals with insentient materials in biological research was the original recommendation made by Russell and Burch (1959) in their influential book on the 3R approach (refine, reduce, replace). However, the recommendation

to use *relatively* insentient animals that are "low on the phylogenetic scale" has become more widely accepted by the scientific community (Tannenbaum & Bennet, 2015; Franco et al., 2018). Indeed, the scale-based approach to replacement has been codified in the *Guide for the Care and Use of Laboratory Animals* (National Research Council, 2011b), which is used to regulate animal research in the United States; many other countries have similar guidelines. Among the general public, too, the vast majority (80% to 90%) accept the use of animals in medical research so long as animal welfare is optimized and alternatives have been explored (Festing & Wilkinson, 2007). Importantly, there appears to be a broad consensus that research on "lower animals"—notably pests and animals we kill for food—is more acceptable than research on nonhuman primates and animals we like to keep as pets (see chapter 3). This view is what one might call a moderated form of the welfarist's perspective.

This more moderate, scale-based approach to animal replacement can be criticized for its reliance on the notion of a phylogenetic scale, which is inconsistent with how evolution actually works (see chapter 2, section 2.5.1) and is shaped by rather subjective preferences of some species over others (a form of speciesism, although this term is usually restricted to a bias in favor of humans alone) (Singer, 1990; Oberg, 2016). However, even if we reject the concept of a single, linear phylogenetic scale, it does seem reasonable to suppose that the degree of sentience (if we accept that sentience comes in degrees) correlates with nervous system complexity (Bullock, 2002; Mather, 2008) and that, therefore, monkeys or chimpanzees will likely suffer more than, for example, a nematode, a fly, or a zebrafish larva when undergoing analogous experimental procedures. Similarly, most people would agree that cultured cells—even in the form of organoids or organs-on-a-chip (see chapter 4)—are less sentient than vertebrates. If that is true, then why not replace animals that society considers to be highly sentient with less sentient animals or insentient cells? Even if our judgments about degrees of sentience are somewhat subjective (Striedter, 2016), is this not a prudent, righteous strategy?

The problem is that shifting research to less sentient or insentient systems implies a shift toward systems that are more different from humans, at least on average. When it comes to biomedical research, the overall similarity between an animal or in vitro model and the human condition—what Russell and Burch (1959) called the model's "fidelity"—tends to decrease with its position on the phylogenetic scale or, to be more objective, its phylogenetic and genetic distance from us (see section 7.1.2). Similarly, even the most elaborate in vitro systems are a far cry from their in vivo counterparts in terms of overall complexity. This decrease in model fidelity is a problem if we assume that our ability to translate findings from a model system to humans correlates with model fidelity.

That said, successful models do not need to be similar to their target in all respects; it is important only that they mimic the target in the features one is trying to model.

For example, viruses, bacteria, and yeast were extremely useful models in the early days of molecular biology because the phenomena of scientific interest at the time turned out to be broadly conserved. For example, the ability of X-rays to damage DNA could be studied in a wide variety of simple animals and even cultured cells because the underlying mechanisms are very general. In fact, microbes and cultured cells were far more convenient for this kind of research than more complicated animals, and thus were better models in practical terms. As we reviewed in chapter 3, the situation changed only as researchers turned their attention to more complex biological problems that were less broadly shared and, therefore, required research on more complex animals, such as fruit flies. One might even argue that biomedical progress over the last 50 years entailed a steady increase in the complexity of the problems being investigated and that this increase required a shift to ever more complex model systems, notably mammals. At least biologists are now tackling a broader array of problems, which requires a broader array of species for research.

Given all these considerations, using mice as model animals seems like a reasonable compromise between having a relatively high-fidelity model and relatively low ethical barriers. After all, mice are mammals like us, yet widely reviled as pests (Little, 1935). Maybe we can exclude them from our concern for animal welfare (as US legislation does; see chapter 2). Moreover, their rapid rate of reproduction and relative tolerance of inbreeding make mice well suited to strain standardization and genetic analyses (see chapter 3). It makes sense, therefore, that mice have played such a prominent role in biomedical research over the last few decades (Libby, 2015). Another, very different compromise between high model fidelity and low ethical concerns can be achieved by performing experiments on cultured human cells. They are genetically human yet presumably insentient (so long as cultured mini-brains remain quite different from human brains; see chapter 4).

Presented with these two main compromise options, one may debate their relative merits. In practice, however, it is more sensible to embrace a tiered approach in which one begins research with relatively simple, low-fidelity models and then proceeds to progressively "higher" models. The idea is that the simple systems allow researchers to explore and develop hypotheses that can then be tested for generality. This tiered, sequential research strategy is common in toxicology and is often pursued by companies involved in drug development. Moreover, it is no accident that regulatory agencies often require safety and efficacy data from two different species, one of which is not a rodent. Obtaining the nonrodent (often nonhuman primate) data is usually the last step before proceeding to clinical trials.

Many advocates for animal welfare are not in favor of the tiered research strategy, however, because it does involve some higher animals, at least at late stages of

translational research. They would prefer that scientists shift farther away from animal research and instead focus more of their energy (and research funds) on clinical research. Indeed, careful clinical observations have played a major role in drug development—for example, by revealing unexpected beneficial drug effects (see chapter 6). Epidemiological studies have also played critical roles in many different areas of medical research, including cancer biology, the discovery of risk factors for cardiovascular disease, and toxicology (see chapter 5). In addition, genetic analyses of human subjects can lead to the identification of some disease mechanisms. The cause of sickle cell anemia, for example, was greatly clarified when research showed that the disease in humans is associated with abnormally shaped red blood cells and a mutation in the hemoglobin gene (Frenette & Atweh, 2007).

Unfortunately, purely observational studies on humans are limited in what they can tell us about the underlying disease mechanisms. As Lord Moulton—a famous mathematician and great advocate for medical research—supposedly once said, "when we are reduced to observation, Science crawls" (Edsall, 1969, p. 467). Moulton was right, but experimental studies on humans must be carefully monitored and tightly controlled (e.g., Kalm & Semba, 2005). History is replete with incidents of human experimentation that were later regarded as unethical ("Unethical Human Experimentation," 2020). For example, the psychiatrist Robert Heath in the 1950s and 1960s conducted poorly controlled deep brain stimulation experiments in humans without clear medical benefits. In his most notorious experiment, Heath attempted to "cure" a man of homosexuality by stimulating his reward circuitry while the subject engaged in sex with a female prostitute (Oliveria, 2018). Ethical aberrations such as these led the World Health Association in 1964 to draft the Declaration of Helsinki, which was amended in 1975 to state explicitly that "concern for the interests of the subject must always prevail over the interests of science and society" (Carlson et al., 2004, p. 709). Although this declaration continues to be debated and revised (World Medical Association, 2013), it is widely accepted and codified in regulations around the globe.

Even now, however, unethical human experiments remain a matter of concern. For instance, the physician scientist Paolo Macchiarini between 2008 and 2014 transplanted stem cell–covered plastic tracheas into multiple patients without being transparent about his results, which were quite often lethal (Rasko & Power, 2017). As an indicator of how serious this problem was, Macchiarini has now been indicted in Sweden for aggravated assault (Schneider, 2020). Even when conducted with the best of intentions, human experimentation often creates complex ethical dilemmas (Edsall, 1969). In short, shifting research from nonhumans to humans does not cause ethical concerns to disappear entirely.

7.1.2 Animal Nepotists: Favoring Our Relatives

Rosenblueth and Wiener wrote in 1945 that "the best material model for a cat is another, or preferably the same cat" (p. 320). Similarly, one may well argue that the best material model for a human is another human and that researchers should therefore emphasize clinical studies over research on nonhumans. This argument is often made with an eye toward promoting animal welfare (see section 7.1.1). However, as we discussed in chapter 2, the principal motivation behind the use of models in science is that they are more accessible to experimental investigation than the target system; and as we just discussed, human experimentation is perforce limited. So if we cannot experiment with humans, why not use the next best thing: nonhuman primates? We may not want to perform invasive procedures on our closest relatives, the chimpanzees, or even other apes, but why not macaques or marmosets (see chapter 3)? Averaging over all traits, they are bound to be more similar to humans than any nonprimate, simply because we are more closely related to other primates than to nonprimates.

There is some logic to this argument, as similarity does decrease on average with the square root of phylogenetic distance (Letten & Cornwell, 2014; Striedter, 2019). It certainly seems reasonable to say that "in general, species that have diverged most recently have the closest resemblances in DNA sequences and functions of protein and RNA derived from these sequences" (National Research Council, 1985, p. 16). Within primates, the Old World primates (including both macaques and us) are more similar to one another, on average, than they are to the more distantly related New World monkeys (e.g., marmosets). Specifically, the genomewide identity of protein coding sequences is 94% for humans versus macaques and 91.7% for marmosets versus humans (Preuss, 2019). Of course, these observations hold only on average, and some traits do not follow the similarity-distance rule. For example, marmosets, like humans but unlike macaques, form strong male-female bonds and exhibit extensive, prolonged paternal care (Fernandez-Duque et al., 2009).

A more general critique of the nepotist's view would focus on the fact that many of the traits that are shared by all primates are also shared by nonprimates and can, therefore, be modeled in them. It becomes important know, therefore, which traits are phylogenetically conservative (i.e., widely shared) and which evolved more recently, either with the origin of primates or within primates. This is no easy task, because we cannot know for certain which trait is present in which species until we look, a paradox sometimes called the "extrapolator's circle" (Steel, 2007; Bolker, 2009). However, we do know by now that some organ systems evolved more rapidly than others within the primate lineage. Especially the immune and central nervous systems have diverged far more than, say, the cardiovascular and skeletal systems. Moreover, it seems reasonable to assume that traits associated with the more conservative systems have a greater

likelihood of being modeled successfully in nonprimates. At least one might adopt a tiered approach in which nonprimates are studied first, and primates are called upon only if those initial efforts prove unsuccessful. For example, the observation that the human immunodeficiency virus (HIV) and hepatitis C viruses infect only humans and chimpanzees has been used to justify research on the latter species for those two specific diseases (Institute of Medicine & National Research Council, 2011).

We can conclude that, as a general rule, the odds of successful translation from a model to its target increase with phylogenetic relatedness, but it is never guaranteed. Even when we compare humans and chimpanzees, one or the other species may well be divergent in any specific trait.

Another very important caveat to the nepotist's perspective is that rates of evolutionary change tend to vary across lineages because of differences in generation time, DNA repair mechanisms, population size, and various other factors (Thomas et al., 2006, 2010). This variation in evolutionary rates can confound the general relationship between overall similarity and phylogenetic relationship. For example, it is well established that rodents and primates are more closely related to one another than to pigs, but the genomes of humans and pigs are more similar to one another than they are to that of mice (figure 7.1), mainly because mice have a much shorter generation time (and hence a faster rate of evolutionary change) than either humans or pigs. Similarly, it is not the case that humans are more closely related to dogs than to mice, even though the human and dog genomes and proteins are more similar to one another (Thomas et al., 2003; Asher et al., 2009).

Given these considerations, it is tempting to replace the outdated notion of a phylogenetic scale with the notion of a genetic divergence scale, in which mice and rats rank lower than cats, dogs, pigs, cows, and all the nonhuman primates (see figure 2.2). However, to the nepotist's distress, neither phylogenetic divergence nor overall genetic similarity are perfect predictors of how well a given trait will translate from a model species to humans.

7.1.3 The Perspective of a Pragmatic Optimist

I suspect that many biologists would listen to the arguments presented in the previous two sections and conclude that, yes, working with model systems is complicated and does not always translate effectively, but with the flood of recent technical advances success is probably just around the corner. The failures of translation are mainly due to technical limitations, poorly designed preclinical research, and faulty clinical trials. The heavy emphasis on mice and a few other model organisms is not to blame because model selection always depends on the question that is being asked, and biologists usually choose whichever model is most convenient for answering the question at

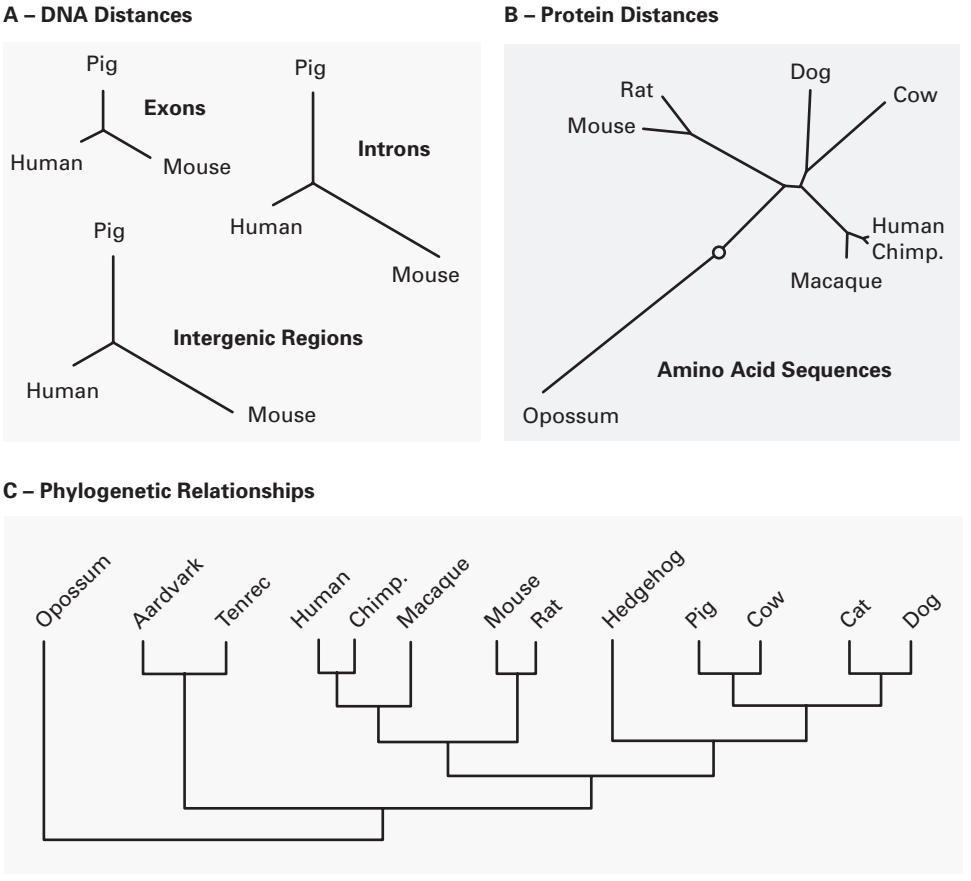


Figure 7.1

Genetic distance versus phylogenetic distance. (*A*) DNA distances. Wernersson et al. (2005) compared the DNA sequences of humans and mice with a large fraction of the pig genome (it had not all been sequenced at the time). The branch lengths in these diagrams represent the degree of estimated sequence divergence and demonstrate that the genetic distance between humans and pigs is less than that between humans and mice. (*B*) Protein distances. An analysis of more species by Cannarozzi et al. (2007) compared protein (rather than DNA) sequences, but again found that rodents are genetically quite divergent from primates. (*C*) Phylogenetic distances. Despite the findings shown in A and B, most biologists agree that humans are more closely related to rodents than to other nonprimates. The reason for the discrepancy between genetic distance and phylogenetic distance (see also figure 2.2) is mainly that rodents reproduce more rapidly than larger mammals and therefore diverge more rapidly. Adapted from (A) Wernersson et al. (2005); (B) Cannarozzi et al. (2007); (C) Asher et al. (2009).

hand. Such optimists are likely to acknowledge that model selection is also influenced by various other concerns, such as researcher training and experience, animal welfare, funding, and the need to produce publications at a reasonable rate. However, they do not think it is imperative to consider species, sex, or other differences between model systems, at least so long as the research is addressing "fundamental" questions about molecular interactions and cellular phenomena underlying disease.

It certainly is true that recent advances in biological techniques have been astonishing. In fact, progress is so rapid that few can keep up, which is one reason why collaborative research has become more important than ever (Wu et al., 2019). Moreover, the new techniques do overcome many of the limitations of the older technology. It is now much easier, for example, to modify specific DNA sequences in a variety of species, generate and differentiate stem cells, and culture three-dimensional tissues. Even some species differences between humans and the models can be overcome by humanizing specific genes or part of the model's immune system (see chapter 3). In light of these advances, some optimism is clearly warranted.

Likewise, it is true that clinical trials can be improved. Clinical trials have already come a long way in terms of being carefully designed, monitored, and analyzed (Junod, 2008; Chow & Liu, 2013). There are, however, ongoing discussions about how heterogeneous the subject pools should be, how far along in the disease process subjects ought to be, how long they should be monitored, and whether tests for efficacy should be performed earlier in the trial pipeline (rather than in the phase III trials). Unfortunately, a major obstacle to many of the proposed changes is increased cost, which is already very high for any late-phase clinical trial. Nonetheless, better enrollment strategies and analytical methods are being developed (Lin & Lee, 2020).

Discussions about improving the rigor and reliability of preclinical research are even more prominent. The topic is too large for us to discuss in depth, but the upshot is that many critics suggest preclinical studies should be more like clinical trials. That is, the sample sizes should be large enough to observe the expected effects (i.e., the studies should be adequately powered), the subjects should be randomly assigned to different treatment groups, proper control groups ought to be included, the data should be analyzed blindly (i.e., without knowing a subject's group assignment), and outcome measures should be specified ahead of time. Ideally, the entire experiment should be registered in advance, and the results should be published regardless of how they turn out. Naturally, such "preclinical trials" (Mogil & Macleod, 2017) would be labor intensive and expensive. They might also discourage exploratory research. Therefore, such rigorous experimental designs should be implemented only for a "final, crucial, confirmatory experiment" (Mogil & Macleod, 2017) that would, if successful, lead to a clinical trial. Although this strategy would be costly and reduce publication

rates for the participating investigators, it would cut down on the extremely frustrating experience of realizing after a failed clinical trial that the preclinical data had been inadequate (Perrin, 2014; O'Collins et al., 2017).

Despite these valid points, blaming insufficient experimental rigor for the profusion of trial failures is problematic because it neglects the fact that successful translation requires replication in a *different* species or system. As Charles Leathers put it in 1990, "the most brilliant design, the most elegant procedures, the purest reagents, along with investigator talent, public money, and animal life are all wasted if the choice of animal is incorrect" (p. 68). It is relatively easy for an investigator to write a sentence at the end of their research paper about how the findings suggest a promising new therapy or treatment approach for a human disease; it is more difficult to know whether the reported results are truly likely to translate or are simply another red herring (Drucker, 2016). To quote Franz van der Staay (2017), "It has hardly ever been questioned in this distending stream of critical reviews addressing lack of translatability, whether the appropriate model animal species have been used" (p. 73). As we discussed in chapter 6, preclinical research may at times suffer from the streetlight effect. That is, the investigators may sometimes be looking for answers in very convenient models that, inconveniently, differ from humans in ways that impede successful translation.

Another potential criticism of the optimist's perspective is that the advantages of novel techniques and therapies are often touted before their limitations become apparent. A classic example is that heroin was originally marketed as a cough suppressant (Sneader, 1998). More recently (and less dramatically) stem cell transplantation showed tremendous initial promise, but using this technique to replace lost cells has proven far more difficult than expected, especially in human brains (Stoker et al., 2017). Similarly, it is distressing to discover that many anticancer drugs currently in clinical trials kill tumor cells in mice even when their putative target molecules are deleted genetically, presumably because the RNA knockdown techniques originally used to identify those putative targets are not as selective as researchers had believed (Lin et al., 2019).

Although this process of diminishing excitement about novel techniques and therapies is natural (part of the "hype cycle"), it becomes a problem when science advances so rapidly that the limitations become a mere afterthought and researchers eager to try the latest techniques move rapidly from one set of inflated expectations to the next. Under such conditions, it becomes difficult for the field at large to know (or even to ask) why some of the promises were never fulfilled.

7.1.4 The View from Comparative Biology

Some biologists bemoan the fact that biomedical research has become heavily focused on a small number of species generally referred to as "model organisms." These dozen

or so species (e.g., *Mus musculus*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Escherichia coli*, and *Arabidopsis thaliana*) achieved their special status mainly because they reproduce rapidly, are exceptionally amenable to genetic analyses, and managed to attract a research community that was willing to develop shared resources to facilitate research on that species (Davis, 2004; Leonelli & Ankeny, 2013). Although these model organisms are distributed across multiple plant and animal phyla (Müller & Grossniklaus, 2010), they represent a minute fraction of the total species diversity.

This kind of taxonomic bias is common in many areas of biology (Pawar, 2003; Rosenthal et al., 2017), but it has become especially severe in biomedical research, where mice have nearly displaced most other research animals. Comparative biologists tend to decry this "murine 'model' monotheism" (Libby, 2015) and advocate for more species diversity in biological research, including biomedical research (Bolker, 1995; Preuss, 2000; Manger et al., 2008; Brenowitz & Zakon, 2015; Yartsev, 2017). This goal of studying a more diverse set of species has been made more attainable by the ever-increasing number of sequenced genomes (National Center for Biotechnology Information, 2020) and the recent advances in genome editing, which can be applied to many different species (Hsu et al., 2014).

An obvious risk to increasing species diversity in biological research, however, is that we might sacrifice depth for breadth. Research on a small set of model organism produces detailed knowledge about these systems at multiple levels of analysis (National Research Council, 1985; Ankeny & Leonelli 2020). Distributing the same amount of research effort across a large number of species would necessarily create more shallow knowledge (Schaffner, 1998). To mitigate this problem, comparative biologists tend to accept a compromise that allows for in-depth studies of a few species that then serve as "reference species" for comparative studies (Striedter et al., 2014). The idea is not simply to compare the heavily studied model organisms to one another, but to examine additional species that are near the reference species phylogenetically. One might, for example, study insects other than *Drosophila melanogaster*, ray-finned fishes other than the zebrafish, and rodents other than *Mus musculus*. The benefit of starting with a reference species is that our deep knowledge of that species can guide the comparative inquiry. Knowing what to look for (and how to look for it) makes the comparative research much easier, even if it ultimately reveals some species differences. In general, the comparative research reveals patterns of species similarities and differences that suggest phylogenetic scenarios and provide important clues about which findings are likely to be conserved across which species (Miller et al., 2019). This knowledge, in turn, should facilitate translational success.

Nonetheless, most comparative biologists are less interested in translation itself than in the discovery of general principles (National Research Council, 1985). Although

comparative biology is sometimes viewed dismissively as being analogous to stamp collecting (Lewin, 1982), it has a long, distinguished history of seeking and discovering general biological principles. As long as one knows only about a single species, it is impossible to know how general one's findings are (Beach, 1950). Even when species differences exist, comparative biologists can discover how they represent different manifestations of conserved principles. As Claude Bernard put it in 1865, "The problem of science will consist precisely in this, to seek the unitary character of physiological and pathological phenomena in the midst of the infinite variety of their particular manifestations" (National Research Council, 1985).

August Krogh, for example, was interested in how the general physicochemical principles of gas exchange were implemented in organisms that lived in very different environments (Krogh, 1941). Later observers tend to emphasize that Krogh liked to study "extreme" organism with special adaptations that made them particularly amenable to experimentation (Krebs, 1975; Pollak, 2014; Green et al., 2018). This is true, but his ulterior aim was the discovery of general principles, not idiosyncrasies. There is nothing wrong with scientific analyses of phenomena that are found only in a few unusual species—they may be inherently fascinating or lead to biologically inspired (e.g., biomimetic) technology (Benyus, 2002)—but it would be wrong to think of comparative biology as being concerned primarily with the study of oddities. A good illustration of this point is Peter Getting's (1988) review of the neural mechanisms that animals use to generate rhythmic activity (e.g., walking, flying, breathing). Although his comparative analysis revealed a variety of mechanisms rather than just one, Getting was relieved to find a set of conserved "building blocks" that were combined in diverse ways.

A major challenge for comparatists who do want to extrapolate their findings across species is that such efforts often depend critically on the details of how the general principles are implemented. The devil, as they say, lies in the details. Even seemingly minor changes in the parameters that govern the interactions between conserved molecules, cells, or organ systems can influence whether a specific compound is toxic or safe, effective or inactive, even if the general principles are conserved. Again, the trial of TGN1412 is a good example: the drug had tested safe in monkeys but was harmful to humans because of a species difference in the expression level of a specific protein (CD28) on a specific type of immune cell (Eastwood et al., 2010). Similarly, chocolate is far more toxic to dogs than to humans, not because dogs lack the enzymes to break down theobromine (the potentially toxic ingredient), but because this compound is degraded and excreted more slowly in dogs (Finlay & Guiton, 2005). Returning to matters of human health, two of the metabolic enzymes that collectively metabolize more than 70% of all marketed drugs differ in complexity, expression levels, and catalytic activity between humans and mice (Gonzalez & Yu, 2006). Such seemingly

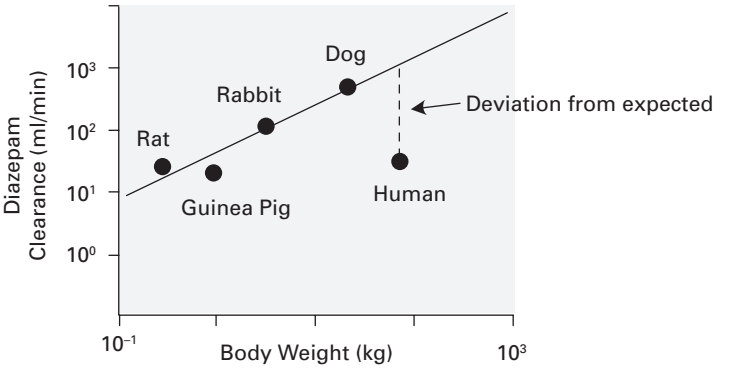
minor, but translationally relevant, species differences are likely to be even more pronounced as we try to extrapolate across more distant relatives (Cross et al., 2011).

One productive way in which comparative biologists have dealt with this variation is to discover "principles of variation" that allow them to predict species differences (rather than similarities) and thus facilitate cross-species extrapolation. For example, they have used comparative neurodevelopmental data to derive a statistical model that can be used to "translate time" across a broad range of mammalian species (Workman et al., 2013). This model helps investigators identify equivalent developmental stages in different species, independently of how rapidly they develop in terms of absolute time (e.g., marsupials develop much more slowly than placental mammals but go through similar stages of development). Such considerations clarify, for example, why thalidomide toxicity tests in diverse species had initially provided confusing results: the drug's effects depend on the developmental stage of the fetus, not on their absolute age (Monamy, 2000). The same general approach reveals the folly of arguing that Alzheimer's model mice do not exhibit the full spectrum of AD symptoms observed in humans because they do not live as long (Foidl & Humpel, 2020): just as dogs age more quickly than humans (T. Wang et al., 2020), so do mice—except more so. As a general rule, life span scales with body size (Speakman, 2005)!

Principles that facilitate cross-species extrapolation have also been discovered in the area of drug dose scaling. To illustrate the problem, consider the fate of Tusko, the adult male elephant who was given 297 mg of LSD in a misguided experiment that killed the elephant within 2 hours of the drug injection (West et al., 1962). The problem was that the investigators calculated the elephant's dose on a milligram per kilogram basis, scaling up linearly from doses known to be safe in humans and cats. This was an inexcusable mistake, for physiologists had long advised that drug doses scale more tightly with metabolic rate than body weight, and that metabolic rate scales against body weight with an exponent of approximately 0.75 (i.e., with negative allometry, rather than linearly). According to the metabolic scaling rule, Tusko should have received only about 1% to 2% of the delivered dose (Harwood, 1963; Boxenbaum & DiLea, 1995).

Later studies showed that factors other than metabolic rate—especially the rate at which a drug is cleared from the body—scale with a variety of different exponents. Therefore, many researchers nowadays obtain data on a drug's clearance rate and several other parameters (notably drug absorption, distribution, and excretion) from multiple species, graph them against body weight, and then use best-fit lines to calculate the predicted parameters for their species and drug of interest (figure 7.2). The predicted parameters are then used to calculate the best-guess dose for administration. In practice, different authors use somewhat different procedures to make their predictions, adjusting them for different classes of animals (Caldwell, 1981; Mahmood, 2007; Sinha et al., 2008). Even

A – Allometric Scaling of Drug Clearance Rates



B – Drugs that Deviate from Allometric Predictions

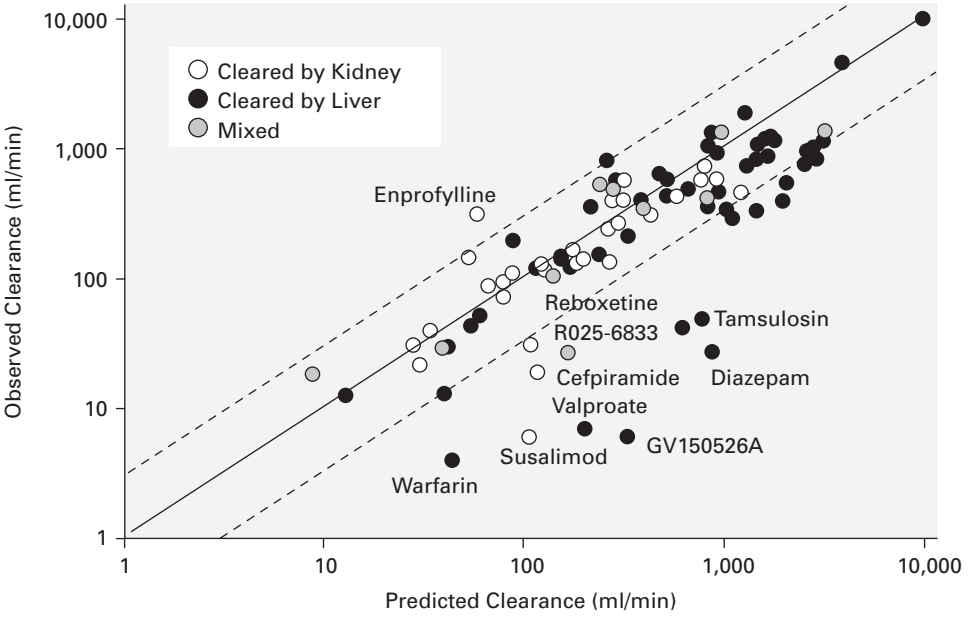


Figure 7.2

Predicting drug clearance with allometry. (*A*) Allometric scaling of drug clearance rates. Extrapolating drug doses across species frequently involves allometric analyses, in which relevant parameters (such as drug clearance rates) are determined in several species. Those data are then graphed against body size and subjected to a linear regression analysis (usually in log-log plots). The resultant best-fit line, in turn, can be used to predict the examined parameter in other species, based on their body weight. This technique works well for many drugs and species, but, as shown here, the clearance rate of diazepam (aka valium) is unusually low in humans. (*B*) Drugs that deviate from allometric predictions. The bottom graph shows the allometric prediction errors for 102 drugs whose clearance rates have been examined in humans and at least three other species (most often dogs and rats). Adapted from (A) Boxenbaum & DiLea (1995); (B) Huang & Riviere (2014), based on data in Tang & Mayersohn (2006).

so, exceptions abound (Aitken, 1983; Calabrese, 1991). For example, the high sensitivity of cats to aspirin defies the predictive procedures, because cats lack an enzyme that is critical to aspirin degradation (Sharma & McNeill, 2009). Still, it is undeniable that the more sophisticated scaling procedures predict safe doses more effectively than the linear extrapolation method that doomed Tusko (Hunter, 2010). Moreover, knowing the scaling rules makes it much easier to identify the "true outliers" (figure 7.2). In aggregate, these comparative studies indicate that data from macaques predict human responses to drugs more faithfully than data from dogs, rabbits, or guinea pigs (Caldwell, 1981), an observation that might please the animal nepotists described in section 7.1.2.

Another medically relevant principle of variation involves cancer susceptibility. A comparative analysis of diverse human tissues has shown that the risk of cells becoming cancerous increases with the number of times their progenitors divided during development (Tomasetti & Vogelstein, 2015). This explains why epithelial tumors, for example, are more common than neuronal tumors (because human neurons tend not to divide after birth). Given this consideration, one would expect large animals with long life spans to develop cancer at a much higher rate than smaller, short-lived species (Seluanov et al., 2018). However, cancer rates do not scale this way. For example, humans undergo about 10^5 more cell divisions in a lifetime than mice, yet the cancer rates of humans and mice are similar. In the words of Rangarajan and Weinberg (2003, p. 952), "about 30% of laboratory rodents have cancer by the end of their 2–3 year lifespan and about 30% of people have cancer by the end of their 70 to 80 year lifespan."

The answer to this puzzle, which is called Peto's paradox (Peto, 1977, 2016), is that humans and other large, long-lived species have evolved various mechanisms that reduce cancer risk (Holliday, 1996; Caulin et al., 2015; Nunney et al., 2015). For example, the number of mutations required to make cells cancerous is larger in humans than in mice (Hahn & Weinberg, 2002). In addition, humans have shorter telomeres and reduced telomerase activity (Hornsby, 2007; Vanhooren & Libert, 2013). Thus, recognition of Peto's paradox predicted and now explains the existence of some species differences in cellular and molecular mechanisms.

7.1.5 A Dialectical Approach: Big Tent Biology

Although the four perspectives I just presented are very different from one another, they are not mutually exclusive. I, for one, oscillate among them. Ultimately, each person—at least each biomedical researcher—must somehow balance the competing interests and attitudes that these perspectives embody. As we pursue this balance, it is important to have access to a broad range of accurate information and to keep in mind that societal attitudes toward the use of models in biology have shifted repeatedly over the course of history—and may shift again.

For me personally, the biggest challenge is to balance my concerns about animal welfare with the desire to alleviate the suffering of sick humans (as well as that of animals in veterinary care). Although an animal or cellular model can be useful even if it does not mimic its target system perfectly, it seems fair to say that successful translation is more likely when the model resembles its targets in as many respects as feasible. It is unfortunate, therefore, that the animals most similar to us (on average) also warrant the greatest ethical concerns. The dilemma was nicely summarized by Karen Rader (2004): "What animals are enough like us to make laboratory results obtained from them generalizable to humans, but not so much like us that we ethically prohibit their being the subjects of experiments?" (p. 22).

There is not one correct answer to this question, as it depends on the research problem under investigation and the stage of the research program. Therefore, I advocate for the use of multiple species in biomedical research, as well as the use of in vitro systems. In effect, I argue for a big-tent biology in which concentrated research on a few reference species is complemented by comparative studies (Striedter et al., 2014), and in which translational research coexists harmoniously with studies that, at first blush, lack direct relevance to human health (i.e., basic biology; Zoghbi, 2013). This inclusive perspective may seem designed to please as many stakeholders as possible (except presumably the anti-vivisectionists), but that is not my intention. Instead, my aim is to encourage every reader—especially the junior scientists (Yartsev, 2017) and shapers of research policy—to think deeply about the relevant issues and then decide for themselves which questions they want to address, which models they consider most promising, and what other kinds of research should be supported.

As we contemplate these questions, we drift toward philosophy. Although biologists generally express little interest in formal philosophy, "scientists in search of answers to real-world issues . . . make decisions constantly that are based on a 'philosophical' stance as to how to do science. . . . For better or for worse, some kind of philosophy is an integral part of the doing of biology" (Orzack, 2012, p. 170). Ideally, those philosophical stances should be developed consciously and, if possible, stated explicitly. To that end, I offer a few recommendations (for an analogous list, see Bolker, 2017).

7.2 RECOMMENDATIONS

The selection of a model system is one of the most influential decisions biologists make in their research, and there are many options. The first step should be to identify a large question of interest. The challenge then becomes finding one or more models that are optimal for addressing this research question. The task requires a careful assessment of

one's own and societal attitudes toward animal research as well as a variety of practical considerations, such as the availability of appropriate animal care or cell culture facilities, the feasibility of the required experiments (given the tools available), and the amount of time and resources one can devote to the research. In addition, it is important to learn as much as possible about the strengths and weaknesses of the available model systems and to imagine which new systems might be superior.

Swamped by all these considerations, the overarching question sometimes gets lost. The long-term goals tend to recede as the research homes in on specific puzzles posed by earlier research. This narrowing of focus is understandable and often highly productive, but must questions posed by previous research in a particular model system be answered in the same system? Might a different system offer a better chance of yielding an answer? What would be the costs and benefits of switching to a different model, or adding one? As one contemplates such questions, some points are good to keep in mind.

7.2.1 Know Your Animals, Know Your Cells!

Biologists nowadays tend to order their research animals or cells from a vendor, much like laboratory supplies. In fact, animals and cells are often listed under "materials and supplies" in grant applications. This attitude is ethically problematic, at least when dealing with presumably sentient animals. Moreover, even when the ethical concerns are minimal, the view of animals and cells as research supplies reduces the likelihood that the researchers will inform themselves thoroughly about the peculiarities of their research subjects. Those peculiarities, in turn, can have significant effects on the results of an experiment, especially if they remain unknown and uncontrolled.

In short, biologists can benefit from learning as much as possible about their research animals and cells. It is important to know, for example, that mice are not small rats. These two taxa evolved separately for about 20 million years and diverged considerably in both genome and behavior (Gibbs et al., 2004; Ellenbroek & Youn, 2016). For example, male mice of many strains (including wild mice) are more likely than male rats to fight with other males, which has implications for how to house them in the laboratory (Crawley, 2007; Kondrakiewicz et al., 2019). Another important difference is that rats swim frequently in nature, whereas mice tend to avoid the water; this probably explains why mice are not as good as rats on spatial memory tasks that require swimming (Whishaw & Tomie, 1996). Many other examples of potentially important species differences are mentioned in the preceding pages.

Within a species, strain differences can also be significant. Particularly obvious is that laboratory strains are very different from their wild relatives. For one thing, the laboratory animals tend to be larger and less aggressive. For another, they exhibit numerous differences in physiology, behavior, and gene expression in the brain (Chalfin

et al., 2014). The many different strains of laboratory mice also differ from one another (Pugh et al., 2004; Crawley, 2007). Some become blind or deaf as they mature, which researchers should know before they subject these animals to tests requiring those sensory modalities (Brown, 2007). Others vary in their responses to stress (Van Bogaert et al., 2006; Mozhui et al., 2010). Not surprisingly, strain differences can also be dramatic at the genetic level. In the words of Cutler et al. (2007),

> over a hundred regions of a strain's genome may be amplified or deleted in relation to the C57BL/6J reference [mouse strain]. This intra-species variability results in many genes being completely or partially deleted, while many other genes are present at increased copy number in a given mouse strain. . . . This surprising variability in the gene content of inbred mouse strains provides both challenges and opportunities in the use of these strains as models for understanding disease and gene function. (pp. 1743–1744)

The genetic variation across strains becomes especially problematic when mutant lines are maintained on mixed genetic backgrounds, because then genetic tests may be needed to determine which animals are likely to develop which traits (Jankowsky & Zheng, 2017). This may not seem like a significant problem because researchers are typically not focused on strain differences, but the genetic background effects can be quite dramatic. For example, one line of mice carrying a mutant gene linked to Alzheimer's disease was developed on a mixed background, but the animals died prematurely when bred onto the pure background of Black-6 mice (Carlson et al., 1997). Even substrains of the popular Black-6 mice differ in a number of potentially important physiological, biochemical, and behavioral respects (Mekada et al., 2009; Simon et al., 2013). Yet another complication is that mutant model mice may over time diverge genetically from the initial stock to the point where in some cases the animals no longer exhibit the phenotype that originally justified their use as a model (Lutz & Osborne, 2014; the Jackson Laboratory, 2021).

As if all this genetic variation were not troublesome enough, biologists must also worry about the many environmental factors that can significantly modify their research animals. For example, laboratory mice are typically raised in rather barren cages that offer little room for exercise but plenty of monotonous food. Under those conditions, the animals tend to become obese, physiologically compromised, and cognitively impoverished, relative to animals that grow up under more natural, enriched conditions. These differences, in turn, can affect how the animals perform in various tests and respond to treatments such as exercise (Würbel, 2007; Martin et al., 2010; Schellinck et al., 2010). As mentioned in chapter 2 (section 2.2.2), researchers also tend to keep their mice in conditions that are colder than the mice would like (i.e., subthermoneutral conditions), which can influence the outcome of experiments (see

figure 2.2; Kokolus et al., 2013). In addition, rearing conditions can affect an animal's microbiome and modify its immune system (Tao & Reese, 2017; Masopust et al., 2017). Finally, researchers should know that many animals are influenced by their cagemates (Baud et al., 2017) and by the humans who handle them. It is startling to realize, for example, that the performance of mice in some behavioral tasks is influenced by the experimenter's sex (Sorge et al., 2014).

Recognizing that experimental outcomes often depend on a large variety of environmental and genetic factors, good researchers tend to control their research animals and experimental conditions rather tightly, which typically means keeping them highly standardized (Nelson, 2018). This strategy makes it easier to obtain consistent results, even if many of the variation-inducing factors are incompletely understood. Unfortunately, results that were obtained only under very specific conditions are unlikely to generalize to other species—that is, they are unlikely to translate well (see section 7.2.2).

Analogous considerations apply to in vitro research. It is rather alarming, for example, how frequently researchers fail to realize that their cell cultures are overrun by HeLa cells or other rapidly dividing cells, even though reports of such cross-contamination have been published repeatedly since 1968 (Gartler, 1968). One study estimates that as many as 20% of all cell lines are misidentified, and another study lists 360 such lines (Hughes et al., 2007; Capes-Davis et al., 2010). It is also common for researchers to use cells that have multiplied in culture for so long that, unbeknownst to the experimenter, they are likely to have changed substantially from their original state (Kaur & Dufour, 2012; Singer et al., 2017). Most cancer cell lines, for example, have diverged somewhat from the original tumors, and some are "hypermutated"—yet they are commonly used (Domcke et al., 2013). Of course, important discoveries can be made with cells that have evolved divergently under in vitro conditions. HeLa cells, for instance, are highly derived (see chapter 4) but have made innumerable contributions to biology. Still, it is clearly a good idea for biologists to learn as much as they can about the cells that they are working with and then to make a conscious decision about their suitability for answering the questions at hand (Krishna et al., 2014).

As part of that decision process, researchers should contemplate the sex of their cells. An astonishing 75% of in vitro studies and 80% of the rodent cell lines distributed by major cell line repositories fail to specify the sex of the cultured cells (Shah et al., 2014; Park et al., 2015). Yet male cells differ from female cells in how they respond to toxins, for example (Nunes et al., 2014). As long as sex remains so commonly unspecified, discovering such differences is difficult. In addition, researchers should consider that homologous cells from different species are likely to differ in diverse ways (Yu & Thomson, 2008; Yang et al., 2011). For example, as mentioned previously, "there are several important differences in the hardwiring of the growth-controlling

circuitry of human and mouse cells. . . . Although humans and mice share a common set of protein components, the regulation of their function is distinct enough to generate quite different rules governing their transformation [into tumor cells]" (Hahn & Weinberg, 2002, p. 337).

Induced pluripotent stem cells (iPSCs) derived from human subjects avoid the species difference problem, but they are nonetheless variable and hence difficult to "know." Even when derived from the same individual, induced stem cell lines may vary because of differences in the genome or epigenome of the original cells, because of random changes introduced during the pluripotency induction process, because of mutations incurred while the cells were dividing in vitro, because of differences in the protocol used to differentiate the cells, and because of how far along the cells are in the differentiation process (Carey et al., 2011; Liang & Zhang, 2013; Merkle et al., 2017). To appreciate the importance of this variability, ask yourself: Which cells should be used as controls in studies with patient-derived iPSCs? How can researchers separate the disease-linked features of their cells from other, potentially confounding variables? In the case of Huntington's disease, where the disease-causing mutation is readily identified, control iPSCs can be created by "editing out" the mutation (An et al., 2012). Unfortunately, this is not possible for most other diseases, where the genetic basis is less well defined.

7.2.2 Standardize, but Not Too Much!

Just as the animals and cells used in research vary in diverse ways, so do the methods used to study them. Collectively, all this variation makes it difficult to replicate results, both within a laboratory and across laboratories (see chapter 1). As mentioned in the previous section, scientists often try to combat this problem by calling for stricter standardization of the research subjects and testing methodology (Global Biological Standards Institute, 2013; Freedman & Inglese, 2014). Historically, the extensive inbreeding of laboratory rats, mice, and flies was clearly motivated by a desire to standardize the animals (see chapter 3); similarly, a major motivation behind the creation of immortal stem cell lines was increased standardization of the cells. Researchers also strive to standardize the environmental conditions in which their animals are raised or their cells are maintained, and they try to standardize experimental assays and tests. Yet, despite strong efforts and extensive collaboration between laboratories, variability persists (Crabbe et al., 1999; Wahlsten et al., 2003; Egan et al., 2016; Kafkafi et al., 2018) (figure 7.3). Not everyone agrees that this is bad.

Although increased standardization can certainly be profitable, it does have some downsides. For one thing, premature standardization can leave investigators stuck with inferior systems and methods (Kalueff et al., 2007). For another, extensive standardization

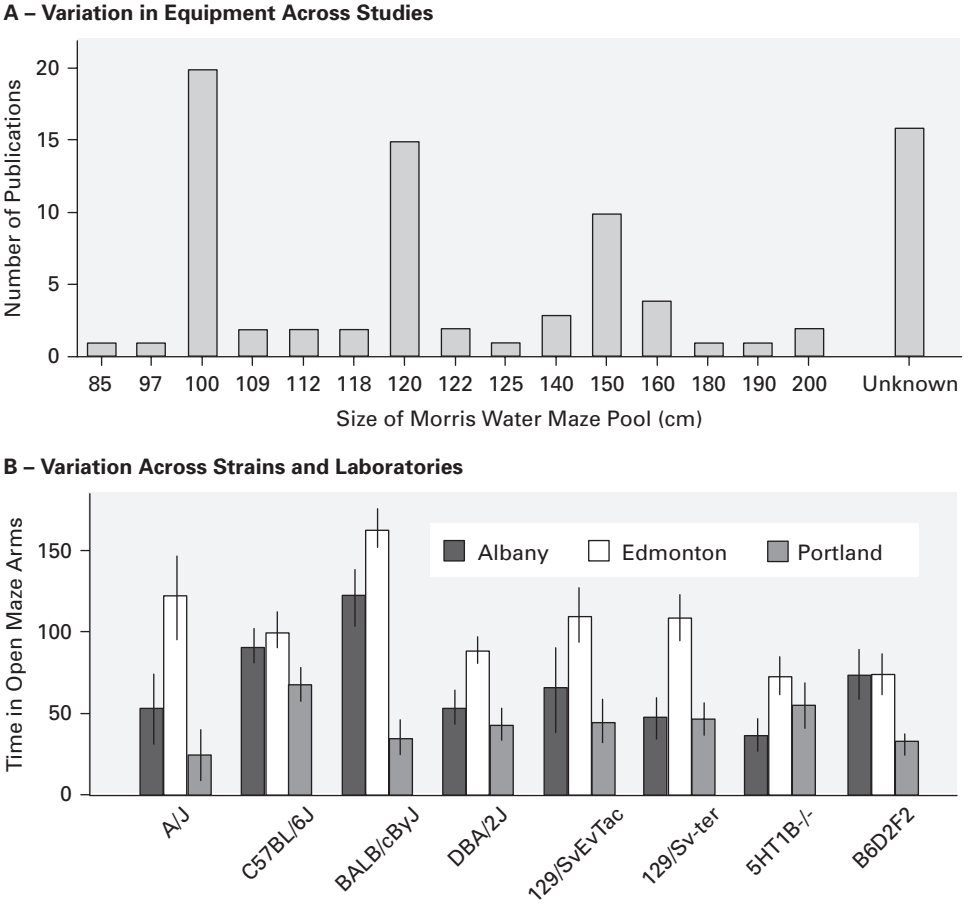


Figure 7.3

Variation across laboratories and mouse strains. (*A*) Variation in equipment across studies. Egan et al. (2016) found that the diameter of the pools used in Morris Water Maze memory experiments performed with Alzheimer's disease model mice varied significantly across 83 published studies (or was not reported). (*B*) Variation across strains and laboratories. Crabbe et al. (1999) compared how mice of different strains and in different laboratories (located in Edmonton, Portland, and Albany) performed on an elevated plus maze with two open and two closed arms. The illustrated measure is the amount of time spent in the open arms, which is a measure of "anxiety" as it is commonly measured in mice; the error bars represent standard errors. Evidently, the variation is substantial, both across strains and across laboratories, despite extensive efforts to standardize experimental procedures. Adapted from (A) Egan et al. (2016); (B) Crabbe et al. (1999).

can lead to the reporting of very small effects that are unlikely to be of use in the real world. For example, highly standardized tumor transplantation studies in mice (see chapter 5) can reveal minute differences in tumor size or growth, but in human clinical trials tumor shrinkage must be on the order of 30% for the treatment to be considered a success (Gould et al., 2015).

A more insidious drawback of standardization is that removing as much variation as possible from an experiment leaves open the possibility that the results may hold only under those specific conditions, which would reduce replicability (Richter et al., 2010). In the words of the famous statistician Ronald Fisher (1953),

> The exact standardization of experimental conditions, which is often thoughtlessly advocated as a panacea, always carries with it the disadvantage that a highly standardized experiment supplies direct information only in respect of the narrow range achieved by standardization. Standardization, therefore, weakens rather than strengthens our ground for inferring a like result, when, as is invariably the case in practice, these conditions are somewhat varied. (p. 99)

The only way to escape this "standardization fallacy" (Würbel, 2000) is to include some variability in the initial experiments or to test for generality as a second step, after having shown significance in a first round of highly standardized experiments. The former approach is facilitated by the adoption of experimental designs and analytical methods that can accommodate variation (e.g., stratified or block designs; Garner, 2014). The latter strategy is implicit in the time-honored (if not always followed) strategy of replicating studies under somewhat different conditions and confirming findings in multiple species before transitioning to clinical trials. The anticancer drug trastuzumab, for example, was shown to be effective in multiple mouse models by several different laboratories before it was tested in humans (Pegram & Ngo, 2006).

One area in which I personally would like to see more standardization is the statistical analysis of similarities and differences across species, strains, or cellular systems. For example, the extent of overlap between lists of differentially expressed genes in two or more systems (e.g., in multiple models of a specific disease) can be evaluated in several different ways (Kuhn et al., 2007; Plaisier et al., 2010; Handler & Haynes, 2019), but all of the approaches entail different sets of assumptions (Lu et al., 2009; Djordjevic et al., 2016; Hargis & Blalock, 2017).

Given this profusion of methodologies, the best approach is probably to compare different models and the target to one another using a single methodology, and then evaluate their relative degrees of similarity. Using such an approach, Burns et al. (2015) determined that some AD model mice are much better than others in terms of modeling human Alzheimer's disease. It is unclear, however, whether different methods of

analysis would yield equivalent results. Similarly, Kuhn et al. (2007) compared the gene expression changes observed in humans with Huntington's disease and several mouse HD models. Although the data clearly indicate some differences between the models (especially in the time course of symptom development), the authors focused on their similarities. As in the controversy surrounding mouse models of human sepsis (see chapter 5, section 5.1.4), the lack of generally accepted standards for comparative transcriptome analyses increases the likelihood that some biologists will see significant similarities where others see great differences.

7.2.4 Learn from Clinical Trial Failures!

Scientific models, both abstract and material (see chapter 2), can promote discovery by revealing phenomena that are evident in the model but not yet known for the target. Thus, a useful model generates predictions that can be tested on the target system. This benefit is widely recognized, but models can also be useful when their predictions are not borne out, when the extrapolation fails. In that case, researchers must reevaluate the model's assumptions and go back to the drawing board.

In biomedical research, the process of learning from failed extrapolations is a form of reverse translation, although this term is often used more generally to describe situations where patient data inspire the creation of disease models (Malkesman et al., 2009; Nadeau & Auwerx, 2019). Despite the general interest in reverse translation, the most common response to failed clinical trials is to start additional trials (with the same or other drugs) or to withdraw from such research entirely. Cook et al. (2014), for instance, report that a major pharmaceutical company used to accept that a certain percentage of its clinical trials would fail and attempted to address the problem simply by conducting a greater number of trials, hoping for a few blockbuster drugs. When this approach did not succeed, the company embarked on a more systematic evaluation of its successes and failures.

Although post hoc analyses of clinical trial failures are rarely published, a few failures did receive detailed follow-up attention. The drug thalidomide, for example, caused a large number of miscarriages and birth defects because it inhibits blood vessel formation in the developing limbs of embryos. Once this mechanism was understood, researchers hypothesized that thalidomide might be effective against cancer by preventing growing tumors from becoming vascularized (Rehman et al., 2011). Indeed, thalidomide is now used in the treatment of multiple myeloma (Palumbo et al., 2008). Another good example is presented by the anticancer drug gefitinib (Ledford, 2008). A clinical trial that had averaged its results across all enrolled patients showed this drug to be ineffective against lung cancer. However, subsequent studies showed that the drug is effective in a subset of patients who have intact, wild-type versions of the *EGFR*

(epidermal growth factor receptor) gene (Paez et al., 2004; T. J. Lynch et al., 2004). Apparently, patients with mutations in *EGFR* had been common enough in the initial trial to confound the results. Thanks to this insight, gefitinib is now approved for the treatment of lung cancer in patients with wild-type *EGFR*.

The analysis of clinical trial failures is likely to reveal mainly information that is specific to the tested drug. However, closer scrutiny of many different failed trials might reveal more general patterns. Many researchers have concluded, for example, that the clinical trials for neurological disorders may have enrolled patients too late in the progression of their disease for the putative treatments to have the hoped-for effect (see chapter 6). They are now testing this hypothesis by enrolling patients earlier, but identifying who is likely to develop a neurological disorder is not a trivial task. Families or isolated populations who have a high prevalence of a specific disease, such as Huntington's disease, may prove useful in this regard (Castilhos et al., 2016).

Along a very different line, I suspect that the high failure rate for drugs targeting neurodegenerative diseases (see chapter 6) stems, at least in part, from species differences in how human and mouse neurons respond to cellular stressors such as the presence of misfolded proteins. Human neurons might be intrinsically more sensitive, or the human brain's immune system might be less effective at preventing cell death. Unfortunately, the literature so far contains only a few hints consistent with this hypothesis (Smith & Dragunow, 2014; Burns et al., 2015; Espuny-Camacho et al., 2017).

7.2.5 Embrace Diversity!

Historically, many biologists—especially molecular biologists—have treated variation between species as a nuisance to be reduced or ignored (Davis, 2003). While this "diversity denial" (Murray et al., 2016) persists in some corners of biology, the attitude toward variation has begun to shift with the rise of comparative genomics and, more generally, the maturation of molecular biology. An important factor in this shift is that genome editing and other technological advances now make it much easier to manipulate the genomes of many different species and cells (Juntti, 2019; Matthews & Vosshall 2020).

Biologists have also started paying more attention to sex, strain, and cell type differences. Such differences have long been reported, but were then neglected. As Hynds et al. (2018) put it in their review of cancer cell lines, the discovery that cancer cells can change dramatically in cell culture should "signal an end to the era in which researchers informally acknowledge that different strains of cancer cell lines are heterogeneous and unstable over long-term passage, but in practice treat them as though they were clonal entities" (p. 4). Of course, similarities should not be neglected either. As Francis

Bacon, the founding father of the scientific method, observed in 1620, some scientists tend to focus on differences, others on similarities; the trick is not to "fall into excess" (see chapter 2, section 2.6).

In this book, I certainly have emphasized differences over similarities, but I have done so with the intent of countering biology's lingering bias in favor of similarities. In general, I think it is best to view similarities and differences as reciprocally illuminating one another (Lehrman, 1971). When confronted with a trove of unfamiliar information, our natural inclination is to seek patterns amid the chaos. However, once those regularities have become apparent, the deviations from the expected come into focus, setting off another round of searching for patterns amid the exceptions. Thus, it is the tension between similarities and differences, between order and disorder, that causes science to advance (Bohm, 1957). The aim, therefore, should be to seek unity within diversity, but not at the expense of diversity.

In particular, I focus in this book on differences that are likely to be medically relevant. Some biologists have recognized such differences for a long time. For example, the biochemist Efraim Racker wrote in 1954 that

> the steadily growing recognition of the existence of alternate pathways, of qualitative and quantitative differences in enzymatic patterns, of differences in submicroscopic cell structure, permeability and rate of cell division have been quoted in favor of a "disunity in biochemistry." The assessment of those features that are not common to various cells might serve to provide us with a better understanding of the disease process as well as its control. (Friedmann, 2004, p. 57)

Extending this perspective, it becomes apparent that there is a considerable amount of disunity not just in biochemistry but in all of biology. For example, many genes have been lost, duplicated, or created de novo in various lineages. Even the broadly conserved "disease genes" have sometimes diverged substantially (see chapter 2). So have their regulatory regions and many of the other genes and proteins with which those disease genes interact. Consistent with this divergence, 27 out of 120 genes that are essential for survival or reproduction in humans are nonessential in mice (Liao & Zhang, 2008). One such gene is *SOD1* (Dickinson et al., 2016), which probably helps to explain why findings from SOD1 model mice have not translated well to human ALS (see chapter 6). Similarly, humans who possess three copies of the wild-type *APP* (amyloid beta precursor protein) gene develop early onset Alzheimer's disease (associated with Down's syndrome), but this is not the case in mice expressing three copies of their own wild-type APP (Wiseman et al., 2015). Recognizing such variation is interesting, but more important is to learn from it. In the words of Fisher and Bannerman (2019),

> we need to recognize variation and use it as a source of insight. Variation has hitherto been seen as a problem and something that should be diminished and reduced at all costs. . . . We suggest that variation could be an opportunity that may allow us to understand and identify disease mechanisms and risk factors and, at the same time, to elucidate treatment strategies on an individual by individual basis. Embracing and understanding variation may be of great benefit for translation. (p. 11)

One way to gain translationally useful insights from variation is to study species, strains, or individuals that are either highly susceptible or resistant to specific diseases (Green et al., 2018). Naked mole rats, for example, are unusually resistant to cancer, which helps explain why they live up to 30 years, far longer than other small rodents. Detailed studies have revealed a variety of mechanisms that contribute to this cancer resistance, including an increased tendency for cells to stop dividing when they contact other cells (Seluanov et al., 2009; Tian et al., 2013). Some of those anticancer defenses in naked mole rats may ultimately lead to cancer treatments in humans. Similarly, researchers have begun to explore why rats and some strains of mice are much less sensitive than monkeys to MPTP, the neurotoxin that induces Parkinsonian symptoms in primates (see chapter 6) (Giovanni et al., 1994; Smeyne et al., 2005). Another interesting line of medically relevant comparative research is the study of individual differences in the propensity of rats to become addicted to drugs (Deroche-Gamonet et al., 2004). Paralleling humans, some rats are far more likely than others to become compulsive drug users, and researchers have started to decipher which factors might account for those individual differences (Belin et al., 2011).

A very different but also promising line of comparative research is the use of recombinant inbred strains (Churchill et al., 2004; Threadgill et al., 2011; Collaborative Cross Consortium, 2012). For this work biologists have assembled "panels" of many different inbred lines (e.g., of mice or fruit flies) that were created by randomly intercrossing two or more different parental strains. Because each line is inbred, researchers reap all the benefits of working with "standardized" animals. However, the lines in the panel differ from one another in both genotype and phenotype. Crucially, the lines have already been well characterized so that phenotypic variation between the strains can readily be correlated with genotypic variation (i.e., mapped). This procedure is especially useful for the genetic dissection of complex traits that involve the interactions of multiple genes, which are difficult to study with traditional genetics (Flint & Mackay, 2009). It also facilitates the study of gene-environment interactions. From a biomedical perspective, working with recombinant inbred lines is an excellent (albeit labor-intensive) way of determining the molecular mechanisms underlying disease resistance or susceptibility (Souza et al., 2020).

Despite the benefits of studying a wide array of species, it is important to reiterate that not everything one finds in our distant relatives will be broadly conserved. Many people seem to believe that evolution creates complex organisms by merely adding novel features to simpler ancestors, which implies that all the features we find in simple organisms should be conserved in more complex creatures. However, evolution often modifies old features, even if they are "fundamental" (i.e., involved in an essential cellular or physiological process). The molecular interactions that control the cell cycle, for example, have diverged substantially between plants, yeast, and multicellular animals (see Striedter, 2019). Similarly, the genome of *Drosophila* features only four chromosomes and 14,000 protein coding genes, but it is more complex than originally thought (e.g., as a result of extensive alternative splicing) and has diverged considerably between the various *Drosophila* species (Holloway et al., 2008; Hales et al., 2015).

Even the two major types of yeast used in research (*Schizosaccharomyces cerevisiae* and *S. pombe*) have diverged to the point where roughly 20% of the genes in either species have no clear homologs in the other (Wood et al., 2002). This rampant divergence between distantly related species (note that the two yeasts have evolved separately for 330–420 million years) does not contradict the notion that similarity tends, on average, to correlate with phylogenetic relatedness. However, it implies that the tiered approach to model systems research is fraught: many of the features we discover in simple, distant relatives may differ substantially from our own. Or they may not. Again, we cannot be sure until we do the comparative analysis.

7.2.6 Reckon with Complexity!

Since the early days of fly genetics (see chapter 3) and the discovery of DNA in 1953 (Watson & Crick, 1953), scientists have tended to embrace a gene-centered view of biology. Although this general approach has undeniably yielded significant progress, it is ripe for reexamination. In particular, we might want to question whether it is appropriate to

> think of each gene as a word in a language, the common language of biology. A word has a specific meaning but it can appear in different contexts. So the word *house* has the same meaning in a very simple sentence that it would have in a complex sentence. Think of genes as the vocabulary of biology. Once this vocabulary is understood, it can be the starting point to understand how genes are assembled to create organisms with a range of different complexities. To learn this vocabulary, we are best off understanding it by reading the simplest text possible. For that reason, we start with something simple like the worm. (Bargmann 2000, p. 520)

This statement is not, strictly speaking, wrong, especially if we emphasize the "starting point" phrase. However, the philosophy of science embedded in this analogy is

problematic for two reasons. First, it disregards the many derived (i.e., not broadly conserved) features of the so-called simple organisms (as we just discussed). Second, it implies that the meanings of words and the functions of genes are largely independent of the context in which they are deployed. This is certainly not true for many words (Gennari et al., 2007). The word "house," for instance, can be a noun or a verb, depending on how it is deployed in a sentence. Genes, too, often have different functions (i.e., meanings) in different molecular and cellular contexts.

Although most genes and proteins are thought to interact with a conserved core set of other molecules, the overall "wiring diagram" of genetic and protein-protein interactions can vary across cell types, cell states, developmental stages, species, and other dimensions (Bandyopadhyay et al., 2010; Ideker & Krogan, 2012; Diss et al., 2013). For example, the transcription factor *Myc* can stimulate epidermal cells either to proliferate or to stop dividing and differentiate, depending on the timing and level of *Myc* expression (Watt et al., 2008).

Gene functions can also vary with the availability of other, interacting genes and proteins. The transcription factor REST (repressor element-1 silencing transcription factor), for instance, targets very different sets of genes in embryonic stem cells versus neural progenitors, most likely because of major differences in chromatin organization between the two cell types (Johnson et al., 2008). Then there are changes in gene sequence that can directly modify a protein's interactions with other molecules. Mutations in the *huntingtin* gene, for example, expand the corresponding protein's interactome (Basu et al., 2013). Major changes in gene function also occur across species, as illustrated by the evolution of the Huntingtin-associated protein Hap1. This protein underwent significant sequence changes in its functional domains that were associated with major changes in the protein's interactome and physiological functions (Lumsden et al., 2016). Another interesting example of evolutionary change in gene function is that deletion of the *retinoblastoma* (*Rb*) gene causes retinal degeneration in humans but has the same effect in mice only if it is deleted in conjunction with an additional tumor repressor gene (Robanus-Maandag et al., 1998; Hahn & Weinberg, 2002). This finding implies some significant genetic rewiring between humans and mice.

One could list many additional examples of genes that have somewhat different functions in different cell types, at different developmental stages, in different species, and so forth, but no matter how long the list, the examples would likely be dismissed (by some) as being exceptions to the rule. Indeed, there is surprisingly little hard evidence to provide empirical estimates of genetic conservation versus change. However, some relevant findings have emerged from studies that use abstract, mathematical models to understand networks of interacting genes and proteins. Specifically, scientists have modeled several well-defined molecular systems, such as insulin or growth

factor signaling, with a set of differential equations (Sedaghat et al., 2002; Brown et al., 2004). Such models typically contain many "free" parameters that are not known but could affect the system's behavior.

A clever way around this problem is to systematically explore the entire parameter space (i.e., test many randomly selected sets of parameters) and ask how differences in specific parameters affect the system's behavior. Such studies have revealed that most of these biological systems are subject to "sloppy control," which means that a relatively small number of parameters are disproportionately influential (Brown et al., 2004; Gutenkunst et al., 2007). The flip side is that many parameters exert little influence over the system's behavior and, therefore, do not need to be empirically determined if one wants to model the system mathematically.

Even more interesting, for present purposes, is that the importance of changes in any one parameter tends to depend on the constellation of the other parameters (figure 7.4). That is, a parameter that is influential within one randomly selected set of parameters may well be negligible in the context of a different parameter set. Moreover, when such parameter sets are subjected to simulated (i.e., in silico) evolution, parameters that exerted great control over the system's behavior in one generation may become unimportant in a later generation, or vice versa. This phenomenon, called "causal drift" (Wagner, 2015), is consistent with the finding that essential genes in one species may be nonessential in another (see section 7.2.5) and that gene knockout effects often depend on the mutation's genetic background (Gerlai, 1996).

More generally, the computational modeling suggests that context-dependent gene and protein functions might be the rule rather than the exception, at least for biological networks of modest size (which is all we can currently model). In the context of translational research, this means that "any one disease determinant that is crucial in one [genetic] background will be modestly important in another and virtually irrelevant in yet another background. In an evolving population that explores the parameter space of such a circuit through DNA mutations, genetic determinants of disease can vary randomly over time" (Wagner, 2015, pp. 2–3).

In other words, we should not be surprised that different species, sexes, and strains sometimes respond differently to different drug treatments, disease triggers, or other manipulations, even when the underlying principles (i.e., the applicable differential equations) remain conserved. It is a consequence of biological complexity!

How should this insight influence the course and conduct of translational research? I do not know, but I submit that biologists are already on the right track insofar as they are paying more attention to both biological complexity and variability (especially sex differences; see chapter 1). In terms of drug development, I agree with Roth et al. (2004) and others (Scannell et al., 2012; Reddy & Zhang, 2013) that we might benefit

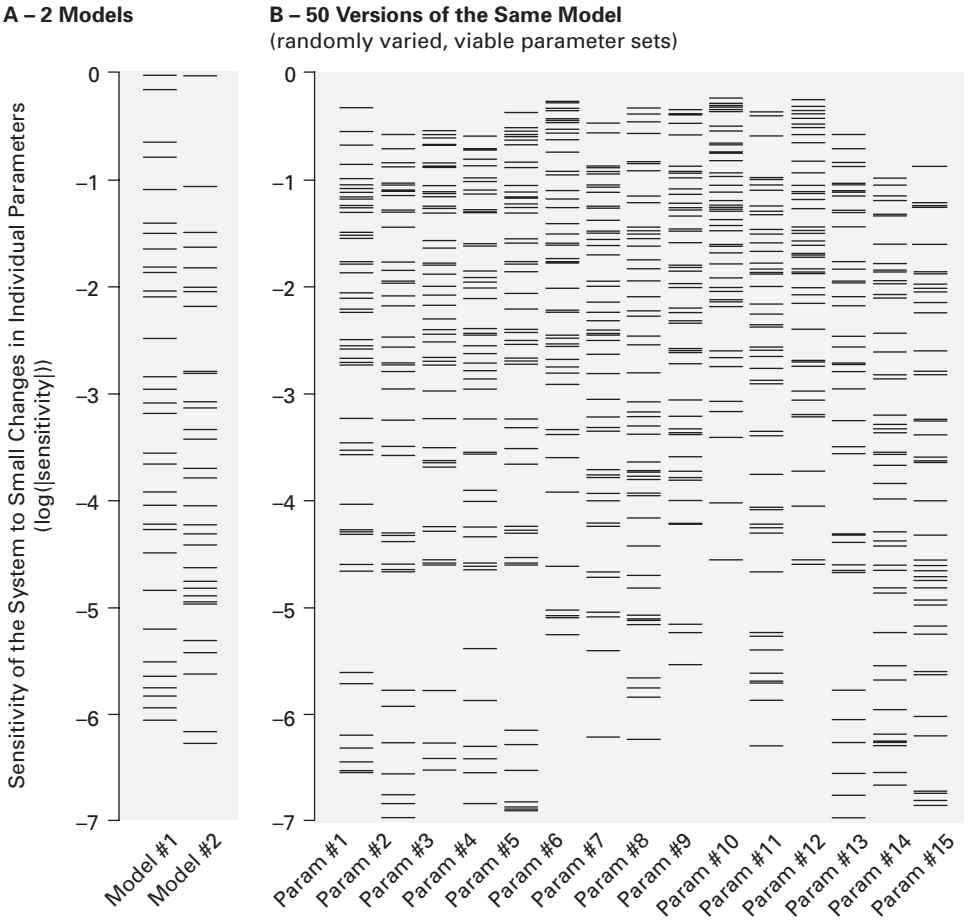


Figure 7.4

Sloppy control and context-dependence in complex systems. (*A*) Two models. Gutenkunst et al. (2007) performed a sensitivity analysis for two biological systems that have been modeled quantitatively—namely, rat growth factor signaling and *Drosophila* segment polarity development. Each horizontal bar indicates how sensitive the model is to small changes in a specific parameter (the closer to zero the plotted data point, the greater the sensitivity). For both models, some parameters have a lot of influence over the system, whereas others exert very little control; this is called "sloppy control" (Brown et al., 2004). (*B*) Fifty versions of the same model. Wagner (2015) performed a similar analysis on a model of cellular insulin signaling, but he did this for multiple, randomly selected, viable parameter sets. Each column shows the sensitivity distribution of a specific model parameter across the 50 randomly generated iterations of the model. Across iterations, the individual parameters varied widely in how much control they exert over the system. The general conclusion is that the importance of any parameter within a complex system depends heavily on the system's other parameters. Adapted from (A) Gutenkunst et al. (2007); (B) Wagner (2015).

from shifting our search image from the highly selective magic bullets promoted by Paul Ehrlich and others (Strebhardt & Ullrich, 2008) to magic shotguns that affect multiple processes, at least for complex diseases. Such multitarget drugs may be more difficult to discover in the laboratory, but history suggests that some fungi, plants, or animals may have already "discovered" compounds with the requisite properties naturally, though natural selection (e.g., aspirin from willow tree bark). I also believe that we have much to gain from further efforts to fight disease by strengthening the body's natural defenses, as in the case of cancer immune checkpoint therapy (see chapter 5). Improving our ability to manipulate the brain's immune system (mainly microglia) would be especially helpful in the fight against neurological disorders.

As biologists proceed with this hard work, I hope that they will increasingly adopt an organismal and comparative perspective, rather than the gene-centered view that has dominated biomedical research for the last 70 years. In essence, I would love to see a merger of the comparative physiological perspective that August Krogh (1929) and other early physiologists embraced (see chapter 2, section 2.6) with the insights and techniques of modern molecular biology. Whether the emerging field of systems biology will expand to incorporate this broader perspective remains to be seen (Joyner, 2011).

7.3 CONCLUSION

My overarching aim in this book has been to increase awareness of the issues that surround the selection and use of model systems in biomedical research. As announced in chapter 1, my focus has been on biomedical research, rather than biology in general, because the high rate of translational failures in biomedicine represents an urgent problem that warrants more debate and analysis. As stated in this chapter's opening quotation, "continuing to do the same things in the same way cannot go on" (Stein, 2015, p. 1259). The question of how to change research conduct or policy is difficult, of course, and I do not have definitive answers. Most of my recommendations are offered not with a sense of certainty, but with the conviction that, after such extensive study of the topic, I really ought to offer at least a few suggestions. Even if my efforts simply prompt some readers to give the topic more explicit thought, I would be pleased.

Libby (2015) and others have pointed out that many of our disease models are not really models as much as they are tools biologists can use to gain pathophysiological insights. I agree with this assessment, but describing animals and cells as "tools" downplays the fact that they have evolutionary and developmental histories that make them different from one another. We can use genetic techniques and environmental

manipulations to alter our models, but we do not build them from scratch; they have their own complexities and idiosyncrasies, many of which remain unknown. Therefore, when we use these material models to gain insights into the human condition, we will sometimes stumble over differences that make a difference. As they say in veterinary medicine, it is "all one medicine until it is different" (Hunter, 2010). Of course, biomedical researchers should try to extrapolate from their models to humans, but they should do so humbly and endeavor to learn from their mistakes. In particular, they should seek principles that can account for the observed variation. They should also be prepared to change their models and, when feasible, use multiple models.

In a stinging critique of biomedical research, David Horrobin (2003) accused the scientists engaged in animal and in vitro research of being unconcerned with real-world clinical problems, but this is unfair. They do care, and most are deeply frustrated when their best efforts at biomedical translation fail. It is simply that there is no easy fix. Still, as in so many aspects of life, becoming more aware of where the problems lie tends to reveal a more productive road forward.

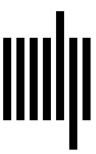
Model Systems in Biology

History, Philosophy, and Practical Concerns

By: Georg Striedter

Citation:
Model Systems in Biology: History, Philosophy, and Practical Concerns
By: Georg Striedter
DOI: 10.7551/mitpress/14366.001.0001
ISBN (electronic): 9780262370028
Publisher: The MIT Press
Published: 2022

The open access edition of this book was made possible by generous funding and support from MIT Press Direct to Open



The MIT Press

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Adobe Garamond Pro and Berthold Akzidenz Grotesk by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Striedter, Georg F., 1962– author.
Title: Model systems in biology : history, philosophy, and practical concerns / Georg Striedter.
Description: Cambridge, Massachusetts : The MIT Press, [2022] | Includes bibliographical references and index.
Identifiers: LCCN 2021033979 | ISBN 9780262046947 (hardcover)
Subjects: LCSH: Animal models in research. | Animal experimentation.
Classification: LCC R853.A53 S77 2022 | DDC 616.02/7—dc23
LC record available at https://lccn.loc.gov/2021033979